Avicennia, 2000, 12/13: 11-18

Aplicación de modelos estocásticos en la evaluación de hipótesis generadas por datos de avistamientos de mamíferos marinos

Application of stocastic models in the hypothesis evaluation generated by watch data of marine mammals

Javier Axis-Arroyo*, Daniel Torruco*, M. A. González* y Benjamín Morales-Vela**.

** Centro de Investigaciones y de Estudios Avanzados del IPN, Unidad Mérida; Carretera antigua a Progreso Km 6, C. P. 97310 A. P. 73 Cordemex, Mérida; Yucatán, México.*
*** El Colegio de la Frontera Sur, Unidad Chetumal. Zona Industrial #2 Carretera Chetumal-Bacalar, C. P. 77000 A. P. 424 Chetumal, Quintana Roo, México.*

Resumen

En el estudio de poblaciones de mamíferos marinos, es común el uso de censos visuales para la clasificación de las zonas de distribución de una especie; la secuencia de los resultados de estos censos en el tiempo, puede ser considerada como un proceso estocástico. Al analizar estos resultados en conjunto con otras variables mediante análisis multivariados, su ordenamiento genera hipótesis donde los avistamientos son variables causales y/o de respuesta a ***n*** variables ecológicas, consecuentemente estas hipótesis pueden ser validadas por modelos estocásticos. El modelo cadenas de Markov se aplicó para validar la hipótesis la cual postula que la distribución espacial del manatí del Caribe depende de las variaciones drásticas en la intensidad del viento. El modelo asignó una probabilidad de 0.40 de que aumente el número de manatíes después de un cambio drástico en la intensidad del viento, de 0.56 de que disminuya y de 0.04 de que no varíe. La validez del modelo dependerá del comportamiento de las variables ecológicas analizadas en el tiempo, siendo válido si las variables analizadas tienen un comportamiento semejante entre años; por lo tanto, la distribución de probabilidades tendrá un valor de significación ***p*** igual al valor de análisis del comportamiento de las variables y el estadístico asociado al valor ***p*** corresponderá al de la prueba de contrastación.

Abstract

In the population study of marine mammals, it's common the censuses visual use for the zones classification of species distribution; the results sequence of these censuses in the time, could be considered like a stocastic process. Upon analyzing these results on the whole with other variables by means of multivariate analysis, their ordination generates hypothesis where they are causal variables and/or answer to ***n*** ecological variables, consequently these hypothesis could be validated for stocastic models. The chains Markov model worked in order to validate the hypothesis which it postulate that the spatial distribution of the caribbean mannate relies on the drastic variations in the wind intensity. The model assigns a probability of 0.40 that it increase the mannate number after drastic change in the intensity of the wind, of 0.56 that it diminish and 0.04 of that it don't vary. The validity of the model will rely on the behavior of the ecological variables analyzed in the time, being been worth if the analyzed variables have a similar behaviour between years; therefore, the probability distribution a significance value of ***p*** same to the value of the variables behaviour analysis and the associate statistic to the ***p*** value will correspond to the contrastation test.

Palabras claves: Cadenas de Markov, Procesos estocásticos, Manatí, Caribe Mexicano.

Key words: Markov's chain, stocastic processes, mannate, mexican Caribbean.

INTRODUCCION

En ecología de comunidades los procesos sinecológicos, aun en las llamadas comunidades clímax, son procesos con un nivel de incertidumbre asociado a ellos, pero que pueden presentar un comportamiento en espacio y tiempo con un cierto grado de regularidad (Gauch, 1982; Jongman *et al.*, 1987), esto permite describir a las variables generadoras del proceso mediante un modelo probabilístico que las considere en forma cuantitativa y no cualitativamente. Cuando las variables que determinan un proceso sinecológico se observan en espacio y/o tiempo fijo y evolucionan en esas mismas dimensiones, el proceso sinecológico es estocástico (Legendre y Legendre, 1984).

En el estudio de poblaciones de mamíferos marinos, es común el uso de censos visuales o de algún índice de abundancia basado en estos para la clasificación de áreas. La presencia de individuos de una especie determinada en un espacio dado en el tiempo, puede ser descrita fácilmente por un modelo probabilístico con variaciones temporales. En este contexto, una secuencia de avistamientos en el tiempo puede ser considerada como un proceso estocástico; si los datos de avistamiento se analizan conjuntamente con otras variables mediante cualquier análisis multivariado, el ordenamiento de estas variables generarían hipótesis sobre procesos en los que los avistamientos serán la variable causal y/o de respuesta, en consecuencia estas hipótesis podrían ser validadas utilizando modelos estocásticos.

En 1996 se analizaron las variables que determinan el uso de hábitat del manatí del Caribe (*Trichechus manatus manatus*) en la costa oeste de la Bahía de Chetumal, Quintana Roo, utilizando el análisis de dirección (Axis-Arroyo, 1998). Este método multivariado, evalúa mediante un modelo generalizado el total de relaciones causales que describen un sistema, pero no las relaciones que se presentan entre las variables. Considerando solo el ajuste de los valores de estas relaciones con la estructura de datos propuesta por el modelo, los valores obtenidos de las relaciones entre variables no tienen robustez estadística por sí solos, pero pueden generar hipótesis sujetas a comprobación.

En el presente trabajo se aplicará el modelo estocástico de cadenas de Markov para evaluar una hipótesis generada por el análisis multivariado, en el que los avistamientos de una especie en un área son la variable de respuesta a ***n*** variables ecológicas.

MATERIALES Y MÉTODOS

Al analizar las relaciones entre las variables, una de las hipótesis obtenidas, fue que la presencia de un manatí en áreas de aguas abiertas (sin resguardo), en dos de las tres temporadas climáticas en la zona (nortes y secas), depende de las variaciones drásticas de la intensidad del viento. Esta hipótesis será la que evaluaremos mediante el modelo estocástico de cadenas de Markov, para lo cual utilizaremos los datos de los muestreos realizados en los días con una intensidad del viento 25% mayor a la intensidad promedio del mes, comparándolos con el muestreo inmediatamente anterior realizado en días sin variaciones drásticas, durante los meses de octubre del 94 a junio del 95; en la Tabla 1 se reportan los datos del número de manatíes avistados en el total de las áreas muestreadas y que servirán como base para el presente análisis. Para el cálculo de las probabilidades de transición los avistamientos entre muestreos se compararon por cada área y no en total.

Un proceso estocástico es un conjunto referido de variables aleatorias {Xt} en donde el subíndice t toma valores de un conjunto T y Xt representa una característica medible en el tiempo t (Gnanadesikan, 1977).

Un proceso markoviano presenta las siguientes características:

a) Es un proceso discreto en el tiempo (t = 0, 1, 2, ..., n). En nuestro caso, está definido por cada uno de los muestreos.

b) Es un proceso con un número finito de estados mutuamente excluyentes. Considerando que el parámetro de análisis es el avistamiento de animales, el proceso tendrá 3 posibles estados en relación al tiempo anterior:

- Mismo número de avistamientos (estado estacionario Pi).
- Aumento en el número de avistamientos (Pj).
- Disminución en el número de avistamientos (Pk).

c) Está gobernado por las probabilidades de transición Pij (el paso del estado i al j está dado por la probabilidad Pij). La transición de estados se determinará por la probabilidad de que se mantenga, aumente o disminuya el número de avistamientos entre los tiempos t y t+1.

d) El estado j del proceso depende solamente del estado i y es independiente de la historia del proceso. El estado en el tiempo i dependerá exclusivamente del número de avistamientos registrados en el tiempo i-1.

Muestreo		Relación entre muestreos	Muestreo		Relación entre muestreos
Anterior	Posterior		Anterior	Posterior	
5	3	Disminución	4	3	Disminución
5	12	Aumento	8	1	Disminución
2	6	Aumento	2	1	Disminución
4	2	Disminución	4	6	Aumento
7	6	Disminución	10	6	Disminución
7	12	Aumento	5	1	Disminución
4	1	Disminución	6	8	Aumento
8	3	Disminución	1	2	Aumento
7	7	Sin cambio	1	3	Aumento
9	4	Disminución	2	3	Aumento
5	3	Disminución	4	1	Disminución

Tabla 1. Número de manatíes avistados en el total de las áreas muestrales, registrados antes y después de variaciones drásticas en la velocidad del viento (>25% de la velocidad promedio de los tres días anteriores).

Table 1. *Mannati number watched in the all areas show them, registered before and after drastic variations in the wind speed (>25% average speed of the three previous days).*

e) Tiene un conjunto de probabilidades iniciales P {X0=i} para toda i. El conjunto de probabilidades iniciales de determinará por la diferencia entre el primer y segundo muestreo.

Al plantear la evaluación de la hipótesis como un proceso Markoviano es conveniente especificar los dos conceptos básicos del comportamiento de un sistema en una cadena de Markov:

a) Estado del sistema: Los valores que describen por completo la posición del sistema en cualquier instante del tiempo son importantes para el comportamiento futuro del sistema; en este caso, el número de avistamientos registrados en cada una de las áreas en cada muestreo; b) La transición de estados (PROTTER Y MORREY, 1969). Al tratarse de tiempos discretos la transición de estados son las variaciones en el número de avistamientos registrados entre los tiempos **t** y $\mathbf{t}_{+1}$, lo cual representará la evolución del sistema en el tiempo.

A cada posible transición del estado **i** al **j** se asocia una probabilidad $\mathbf{P}_{ij}$, que en procesos Markovianos se define como la probabilidad de transición de un paso, con valores $0<\mathbf{P}_{ij}<1$ (tomará valor 0 cuando no sea posible realizar una transición del estado **i** al **j**, y de 1 si estando en el estado **i** pasa exclusivamente al estado **j**), lo cual nos permite definir probabilidades de transición a partir de un estado **i** a un estado **j**. Si tenemos **n** estados obtendríamos una matriz **P** denominada matriz de transición, donde:

$$\mathbf{P} = \begin{bmatrix} \mathbf{P}_i & \mathbf{P}_i \text{ --- } \mathbf{P}_{im} \\ \vdots & \\ \mathbf{P}_m & \mathbf{P}_m \text{ --- } \mathbf{P}_{mm} \end{bmatrix}$$

Considerando los 3 estados del presente análisis, nuestra matriz de transición estaría dada por:

$$P = \begin{bmatrix} \mathbf{P}_{ii} & \mathbf{P}_{ij} & \mathbf{P}_{ik} \\ \mathbf{P}_{ji} & \mathbf{P}_{jj} & \mathbf{P}_{ik} \\ \mathbf{P}_{ki} & \mathbf{P}_{kj} & \mathbf{P}_{kk} \end{bmatrix}$$

Donde:
$\mathbf{P}_{ii}$= No hay cambio en el número de avistamientos después de un estado estacionario.
$\mathbf{P}_{ij}$= Aumento en el número de avistamientos después de un estado estacionario.
$\mathbf{P}_{ik}$= Disminución en el número de avistamientos después de un estado estacionario.
$\mathbf{P}_{jj}$= Aumento en el número de avistamientos después de un aumento.
$\mathbf{P}_{ji}$= No hay cambio en el número de avistamientos después de un aumento.
$\mathbf{P}_{jk}$= Disminución en el número de avistamientos después de un aumento.
$\mathbf{P}_{kk}$= Disminución en el número de avistamientos después de una disminución.
$\mathbf{P}_{ki}$= No hay cambio en el número de avistamientos después de una disminución.
$\mathbf{P}_{kj}$= Aumento en el número de avistamientos después de una disminución.

Esta matriz de transición cumple con las siguientes características del proceso:

a) En cualquier momento cualquier manatí puede estar en cualquier área; por lo que en un área **x** puede observarse cualquiera de los tres estados, es decir los estados son probabilísticos.

b) La probabilidad de que aumente, disminuya o se mantenga el número de manatíes avistados en un área; es relativa al número de avistamientos en el tiempo inmediatamente anterior, es decir cumple con la propiedad Markoviana.

c) La suma de las probabilidades de los 3 eventos de cada renglón de la matriz de transición será 1, es decir tiene a 1 como eigenvalor y ninguno de ellos excede a 1 en valor absoluto.

Considerando los avistamientos entre muestreos por cada una de las áreas, la diferencia entre el primer y segundo muestreo no tendrá un antecesor, por lo que se considerará como el primer eslabón de la cadena markoviana, al no tener antecesor la matriz de transición tendrá solo 3 probabilidades:

$$P = [P_i \; P_j \; P_k]$$

La matriz será un vector al que se le denominará vector inicial de distribución $\{\mathbf{X}_0\}$, con componentes no negativos y sumatoria 1 y que tomará los valores de la siguiente matriz:

$$P_i = [0.02 \quad 0.27 \quad 0.71]$$

Las diferencias entre los estados 2° y 3° y subsecuentes nos generaría la matriz de transición asociada al proceso y que en nuestro caso sería:

$$P_T = \begin{bmatrix} 0.00 & 0.37 & 0.63 \\ 0.01 & 0.31 & 0.68 \\ 0.06 & 0.43 & 0.51 \end{bmatrix}$$

Por lo tanto, esta matriz modela las probabilidades de que se observe cualquiera de las 9 transiciones generadas por los 3 estados. Considerando el tiempo, estas probabilidades de transición deben ser estacionarias para poder deducir la probabilidad de que el sistema se encuentre en el estado **j** después de **n** transiciones, a partir del estado inicial **i**, lo cual se denota por $\mathbf{P}_{ij}^{(n)}$, lo cual es una probabilidad de transición de **n** pasos.

Para que el análisis del proceso Markoviano tenga sentido en el tiempo la matriz de transición deberá de cumplir con las siguientes propiedades:

1) Ser irreductible: que sea posible pasar del estado **i** al **j** en algún número finito de transiciones para los estados **i** y **j**, esto es:

$$\mathbf{P}_{ij}^{(n)} > 0 \text{ para algún } (\mathbf{n}) \text{ para toda } \mathbf{ij}$$

2) Ser aperiódica: un estado **i** tiene un periodo **t**>1 si es posible para la cadena estar en el estado **i** únicamente en múltiplos de **t,** es decir:

$$P_{ij}^{(n)} = 0; \text{ siempre que } \mathbf{n} \text{ no sea un múltiplo de } \mathbf{t}$$

Si una cadena markoviana es finita, irreductible y aperiódica es una matriz ergódica y en consecuencia regular (Morrison, 1976).

Cuando una cadena de Markov es regular la distribución probabilística tiene a una distribución límite común conforme **n** aumenta, independientemente del estado inicial en que empezó el proceso y se determina por:

$$\text{límite}_{n®a} P_{ij}^{(n)} = p_j; \; j = 1, 2, ..., n$$

Donde p_j satisface las ecuaciones del estado estable:

$$\Pi_j \geq 0$$

$$\sum_{j=1}^{N} \Pi_j = 1$$

$$\Pi_j = \sum_{j=1}^{N} \Pi_i P_{ij} \; ; \; j = 1, 2, ..., N$$

Esta última fórmula expresada en notación matricial toma la siguiente forma:

$$\Pi = \Pi P \quad 1)$$

La probabilidad p_j es la probabilidad del estado estable y representa las probabilidades de que el sistema se encuentre en cada estado al principio o fin de una transición, después de que haya ocurrido un número lo suficientemente grande de ellas como para lograr una estacionalidad en las probabilidades.

Considerando que:
La **n**-ésima potencia de la matriz **P** se denota como:

$$P_n = \{P_{ij}^{(n)}\}$$

El vector de distribución en el inicio del proceso es dado por:

$$X^{(n)} = [X_1(n), X_2(n), ..., X(n)]$$

Y el vector de distribución en el inicio del proceso es dado por:

$$\mathbf{X}^{(0)} = [\mathbf{X}_1(0), \mathbf{X}_2(0), ..., \mathbf{X}_n(0)]$$

La relación entre los vectores de distribución en el **n**-ésimo periodo y en el inicio del proceso se establecería por la fórmula siguiente:

$$\mathbf{X}^{(n)} = \mathbf{X}_1^{(0)} \mathbf{P}^n$$

En consecuencia, la distribución de probabilidades en el estado límite es determinada por:

$$X^{(\alpha)} = X^{(0)} L^{2)}$$

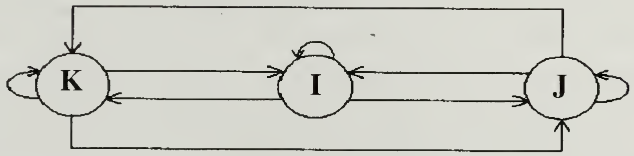
Donde:

$\mathbf{X}^{(\alpha)}$ = Distribución de probabilidades del estado límite.
$\mathbf{X}^{(0)}$ = Distribución de probabilidades en el inicio del proceso.
L = Matriz de transición límite.

En nuestro ejemplo $\mathbf{X}^{(0)}$ toma los valores del vector $\boldsymbol{P}_i$ y la matriz **L** los de la matriz $\boldsymbol{P}_T$.

RESULTADOS

Los estados y posibles transiciones del sistema se especifican en el siguiente diagrama:



En donde:

Los círculos representan los estados.
i = Mismo número de avistamientos.
j = Aumento en el número de avistamientos.
k = Disminución en el número de avistamientos.

Al aplicar la fórmula 2 se obtuvo la distribución límite de probabilidades:

$$X^{(\alpha)} = [0.04 \quad 0.04 \quad 0.56]$$

Lo cual representa que:

Hay una probabilidad 0.40 de que aumente el número de manatíes después de un aumento drástico en la intensidad del viento.

Hay una probabilidad de 0.56 de que disminuya el número de manatíes después de un aumento drástico en la instensidad del viento.

Hay una probabilidad de 0.04 de que no varíe el número de manatíes después de un aumento drástico en la intensidad del viento.

DISCUSIÓN

El supuesto más crítico del proceso markoviano en la presente aplicación (y consideramos que en todas las aplicaciones en ecología), es la condición de la estacionalidad de las probabilidades de transición; esto es, que su matriz P permanezca igual en todo el periodo.

Al considerar tal condición, la validez del modelo dependerá de las diferencias de las variables ecológicas en el tiempo, que determinaron el proceso estocástico. Como en este ensayo se consideraron las temporadas de Nortes y secas, el modelo describirá las posibilidades de transición en tales temporadas y tendrá validez en el tiempo si las variables analizadas tienen un comportamiento estable entre largos periodos.

Estadísticamente el modelo será válido si no existen variaciones significativas entre el comportamiento del total de las variables en los tiempos t y t+1; por lo tanto, la distribución de probabilidades tendrá un valor de significancia p igual al valor del análisis del comportamiento de las variables y el estadístico asociado al valor p corresponderá al de la prueba que contrasta el comportamiento de las variables (Linares et al., 1986).

Si el comportamiento de las variables no tienen diferencias estadísticamente significativas en el tiempo, las distribuciones de probabilidad serán las mismas. Si se presentan diferencias, estas serán un indicador de perturbaciones en la dinámica del sistema y de esta manera, el número de avistamientos tendrá un valor paramétrico que cuantifique las perturbaciones que ocurran en el medio.

Considerando las variables en forma individual, un estado tiene un valor probabilístico generado por un conjunto de datos en los que tÎT; por lo que las variaciones de cada t alteran el proceso estocástico, la estabilidad de las condicionantes causales de estado dependerán del apego de sus distribuciones a la distribución normal; sin embargo, la misma naturaleza estocástica del modelo da un intervalo de tolerancia, por lo que consideramos que una condicionante causal de estado es confiable si los datos de su distribución cumplen con el teorema de Tchebichev.

BIBLIOGRAFÍA

GAUCH, H. G. 1982. *Multivariate analysis in community ecology*. Cambridge University Press. Cambridge.

GNANADESIKAN, R. 1977. *Methods for statistical data analysis of multivariate observations*. Wiley Series in Probability and Mathematical Statistics. John Wiley and Sons. London.

JONGMAN, R.H., C. J. F. TERBRAAK AND O. F. R. VAN TONGEREN. 1987. *Data Analysis in community and Landscape Ecology*. Pudoc Wageningen.

LEGENDRE L., P. LEGENDRE. 1984. *La structure des données écologiques*. Tome 2. 2e. edition. Masson. 335 p.

LINARES-FLEITES, G., L. ACOSTA-RAMIREZ Y V. SISTACHS-VEGA. 1986. *Estadística Multivariada*. Universidad de la Habana. Facultad de Matemática Cibernética. La Habana.

MORRISON, D. F.. 1976. *Multivariate statistical methods*. International Student Edition. McGraw-Hill. Inc. Kosaido Printing Co. Tokyo Japon.

PROTTER, M. H., CH. B. MORREY JR. 1969. *Modern Mathematical Analysis / Análisis matemático*. Ed. Bilingua. Fondo Educativo Interamericano S. A. Bogotá, Colombia.